Sampling Variability and Perceived Threats to Human Health at Great Lakes Beaches

Molly Moynihan, Andrew Gronewold, and Craig Stow
NOAA, Great Lakes Environmental Research Laboratory, Ann Arbor, MI
Collaborators: Vijay Kannappan, and Donna Kashian

Water Quality Standards

The Environmental Protection Agency (EPA) uses microbiological sampling methods to assess water quality at beaches and similar designated use areas, and imposes use restrictions (such as beach closures) based on empirical relationships between concentrations of fecal indicator bacteria (such as *E. coli* and *Enterococci*) and the risk of gastrointestinal illness in swimmers. Bacterial concentrations are monitored at a variety of spatial and temporal scales. Short-term monitoring programs, for example, typically compare water quality samples to Single Sample Maximum (SSM) numeric limits established by the EPA (Table 1). A variety of meteorological and hydrological conditions, however, are known to skew the results of short-term assessments. In addition, conventional analysis procedures have intrinsic variability that is often ignored. Here, we implement a Bayesian statistical analysis to quantify uncertainty and understand how it leads to different perceptions of human health risk and subsequent variability in management decisions.

Designated Use	*E. coli*	*Enterococci*
Designated beach area	235	61
Moderate use - full body contact recreation	298	78
Light use - full body contact recreation	409	107
Infrequent use - full body contact recreation	575	151

Table 1. US EPA Single Sample Maximum (SSM) concentration limits for fresh water. Units are organisms per 100 mL.

Our Approach

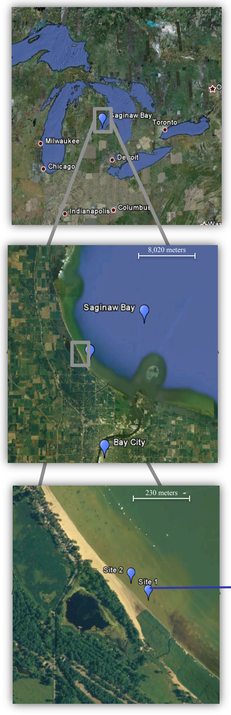


Figure 1. Sampling location - Tobico Beach, MI in Saginaw Bay in western Lake Huron. Sites 1 and 2 were approximately 115 meters apart.

We explore water quality sampling variability by isolating environmental variability from intrinsic laboratory analysis variability and uncertainty in a small-scale spatial and temporal water quality assessment. Data was collected from Tobico Beach in Saginaw Bay in Lake Huron, Michigan (Figures 1 and 2). Samples were collected from 5 different substations at each of 2 sites (Figure 2), and at times intervals of 0, 10, and 20 minutes after the "start" of the experiment. Samples were analyzed for both *E. coli* and *Enterococci* concentration using two different methods: 1) a conventional analysis based on deterministic colony forming unit (CFU) values, and 2) a hierarchical Bayesian approach.

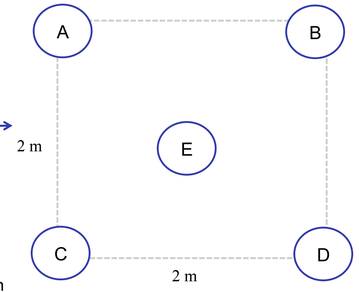


Figure 2. Schematic of sampling grid and positions of substations A-E.

Quantifying Variability

We analyzed each water quality sample in a conventional 4-step process prior to quantifying the bacteria concentration (Figure 3). Step 1 involved preparing different dilutions of the original sample such that at least one of the three dilutions will (in theory) have a "countable" (not zero, and not too high) number of colonies. Step 2 involves filtering each diluted sample. Step 3 – each sample is incubated for 24 hours. Step 4 – the number of colonies on each growth plate are counted. In a conventional analysis, the number of colonies is used, along with the dilution factor, to calculate a CFU value. In our proposed Bayesian analysis, the number of colonies is modeled as a Poisson random variable. The procedure in this figure was repeated for each sampling location for both *E. coli* and *Enterococci*.

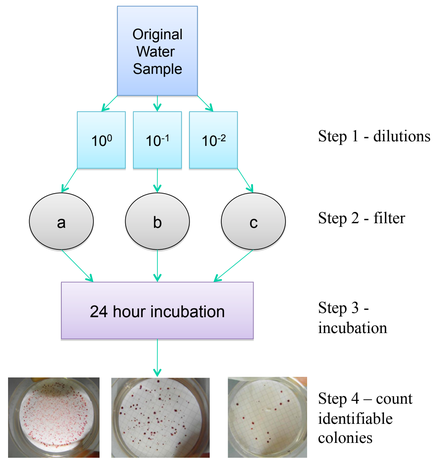
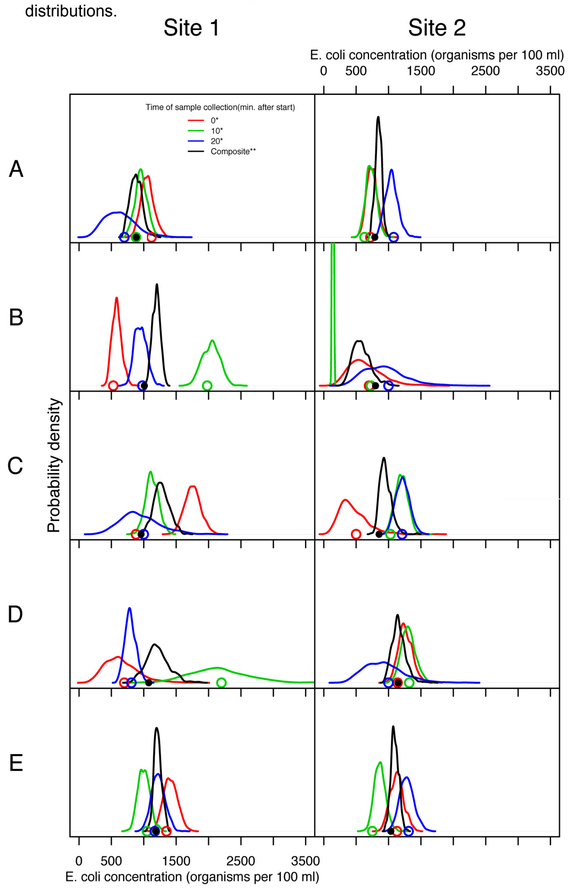


Figure 3. Standard membrane filtration laboratory analysis procedure process schematic.

Figure 4. *E. coli* posterior probability distributions (curved lines) for each sampling station (1 or 2) and substation (A – E) at times 0, 10, and 20 minutes after the start of the procedure (red, green and blue lines, respectively). Black lines represent a composite analysis. Colored dots represent CFU values, following the sample color scheme as the posterior distributions.



Following conventional practice, we combine the number of colonies counted with the dilution factor for each sample to calculate a CFU value (dots in Figure 4). The geometric mean of the three samples at each substation (from times 0, 10, and 20 seconds after the start of the experiment) is then calculated and compared to the single sample maximum (Table 1). We then model the number of colonies counted as a Poisson random variable with mean and variance $\lambda = cV/100$ where c is the "true" but unobserved bacteria concentration (in organisms per 100 mL) in each sample, and V is the volume (in mL) of the original sample. By reversing the logic of this model and using Bayes' theorem, we generate samples from the posterior probability distribution for c for each sample (curved lines Figure 4). We can also combine the information across all samples (from 3 time steps) at a substation to calculate the "composite" concentration probability distribution. This composite probability distribution not only reflects the uncertainty in the laboratory procedures, it also provides a more robust indication of sampling variability than the geometric mean.

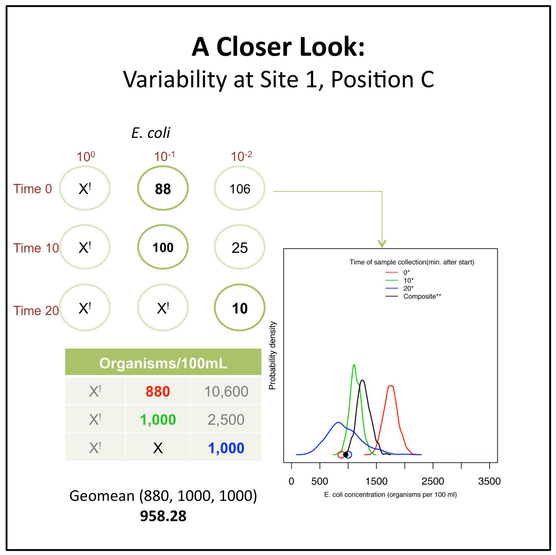


Figure 5. Representative analysis of bacteria samples at site 1, position C. The 9 circles represent growth plates from each of three dilutions across three time steps (times 0, 10, 20 minutes). The table below the circles indicates the CFU value from each sample, and bolded values indicate those which, in a conventional analysis, are used to calculate the geometric mean. The right-hand panel indicates the posterior probability distributions for the *E. coli* at each time step integrated across all three dilutions. Plates with a X! symbol indicate that there were "too many colonies to count", a common reporting procedure in membrane filtration analysis.

Moving Forward: Management Implications

The results of our analysis suggest that the closure status of Tobico Beach could vary quite dramatically depending on minor variations in sampling location and time, and depending on which indicator organism is used as a basis for human health protection and which set of numeric limits is adopted. Using the data from site 1, position C (Figure 5) as an example, the following alternative management decisions could be considered:

Example – issue swimming advisory, or close beach? If Pennsylvania's SSM numeric limits were used as a basis for protecting human health in this assessment, the *E. coli* sample collected at time 0 (with a CFU value of 880 organisms/100 mL) is below the beach closure limit of 1000 organisms/100 mL, but above the swimming advisory limit of 235 organisms/100 mL. Samples collected at times 10 and 20 minutes, however, indicate that the beach closure limit has been met. Including uncertainty in the analysis, and incorporating information from all available dilutions indicates that there is a very significant possibility that the true concentration is well above 1000 organisms per 100 mL.

Previous research has identified discrepancies between water quality standards within the Great Lakes and how they may lead to inconsistent management actions and varying degrees of protection of human health. Our results support these findings, and suggest that a more uniform procedure for monitoring and analyzing water quality needs to be implemented. Monitoring decisions based on a single sample, or even the geometric mean, may result in values unrepresentative of the true water quality. Procedures which explicitly acknowledge uncertainty and variability, such as the Bayesian analysis we present here, allow this uncertainty to be quantified and propagated into a management decision. EPA, however, does not require replicate samples or multiple dilutions. As shown in Figure 5, without replicates, dilutions, and a geometric mean or Bayesian posterior distribution curves, it would be easy to misrepresent the water quality. If this study were repeated by collecting water samples from different depths, conclusions could be drawn as to how fecal indicator bacteria quantities vary throughout the water column on a small spatial and temporal scale.

This work is a product of the multi-disciplinary research of the Center of Excellence for Great Lakes and Human Health, funded by NOAA's Oceans and Human Health Initiative.